User Manual

Zhou Du (adugduzhou@gmail.com)

Version 1.0

1. What is agriGO?

The agriGO is designed to automate the job for experimental biologists to identify enriched Gene Ontology (GO) terms in a list of microarray probe sets or gene identifiers (with or without expression information) and it is also a GO-related database. The agriGO specially focus on agricultural species.

2. Why use agriGO?

The agriGO provides heavy support to agricultural species. Not only limited to SEA analysis, GSEA which is achieved using PAGE method is also available. Furthermore we have BLAST4ID tool for ID transfer or annotation. And search as well as download function is accessible. The agriGO can give out rich outputs like graphical result, bar chart result and hierical tree which composing a comprehensive understanding of biological meaning of user's input data.

3. What data argiGO contains?

We currently support on 35 species including 280 datatypes. Please check the data statistics page for detail information. We will continue adding more species and datatypes.

4. How agriGO prepare its data?

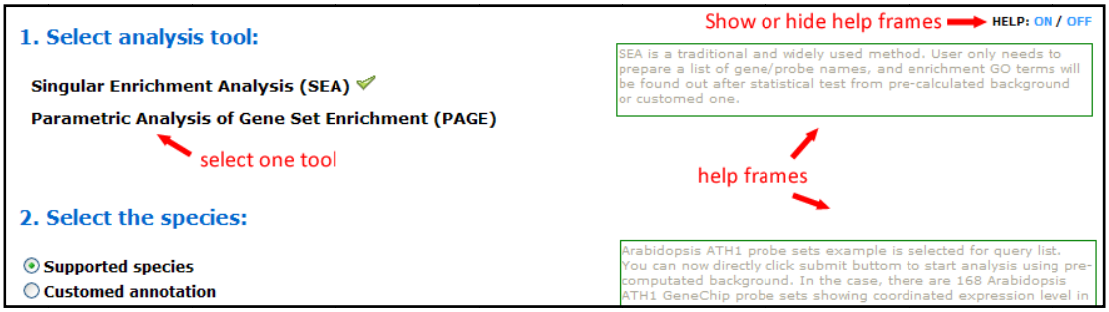
Raw GO annotation data is generated using BLAST, Pfam, InterproScan by agriGO or obtained from B2G-FAR center or from Gene Ontology. Arabidopsis genome data is from TAIR. Rice TIGR genome data is from Rice Genome Annotation Project. Rice KOME data is from KOME database. Rice Gramene data is from Gramene center. Populus genome data is collected from JGI. Soybean and Sorghum genome data is compiled from phytozome. Grape genome data is compiled from Genoscope. Medicago genome data is from Medicago truncatula sequencing resources. Maize genome data is from MaizeSequence.org. Castor bean genome data is from Castor Bean Genome Database. Brachypodium distachyon genome data is from Ensembl. Bovien genome data is from Bovine Genome Database. Silkworm genome data is from SilkDB. M. grisea genome data is from Magnaporthe grisea Database. affymetrixmetrix CSV files and array sequences are from NetAffx.

5. How to use tools provided by agriGO?

Quick introduction to analysis procedure

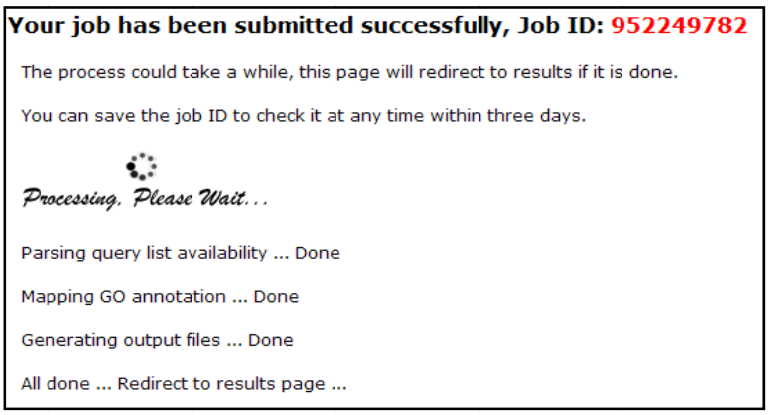
1. Choose tool and set parameters

You should choose one tool to go forward. At the right side, several frames containing annotation text are interactive. The content will change depending on exact parameters you chose. You can make the help frames show or hidden by using HELP buttons at top-right of the page.


Show or hide help frames
HELP: ON / OFF
1. Select analysis tool:
Singular Enrichment Analysis (SEA)
Parametric Analysis of Gene Set Enrichment (PAGE)
select one tool
SEA is a traditional and widely used method. User only needs to prepare a list of gene/probe names, and enrichment GO terms will be found out after statistical test from pre-calculated background or customed one.
help frames
2. Select the species:
Supported species
Customed annotation
Arabidopsis ATH1 probe sets example is selected for query list. You can now directly click submit buttom to start analysis using pre-computated background. In the case, there are 168 Arabidopsis ATH1 GeneChip probe sets showing coordinated expression level in

2. Submit your job and perform analysis

After submitting your job, the agriGO will pre-check the validity of your upload data. If your job is submitted successfully, a job ID will be given. Since the analysis process could take a while, you may close the waiting page and use the job ID to check the work later. Please note that results of your jobs will be stored on our server for **THREE DAYS**. After 3 days all information of the job will be deleted. If you want elongation contact me.


Your job has been submitted successfully, Job ID: 952249782
The process could take a while, this page will redirect to results if it is done.
You can save the job ID to check it at any time within three days.
Processing, Please Wait...
Parsing query list availability ... Done
Mapping GO annotation ... Done
Generating output files ... Done
All done ... Redirect to results page ...

3. Explore results

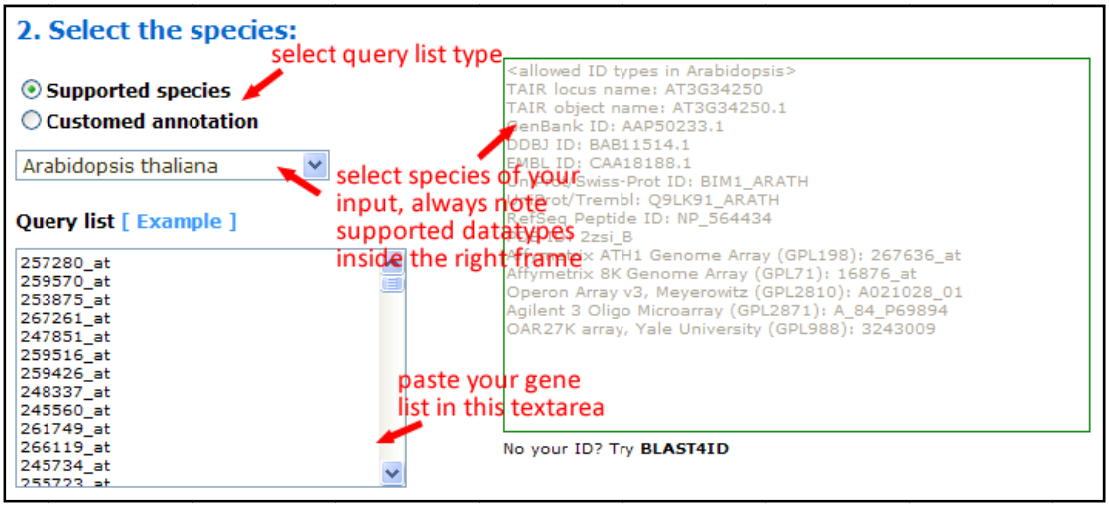
The agriGO provides different ways to browse results of different tools. Some of them are flexible but you may need some specific setting to make them to castor to your own demands. And detailed introduction to these tools in the manual in the following will help you to achieve it.

A. How to use Singular Enrichment Analysis (SEA) analysis?

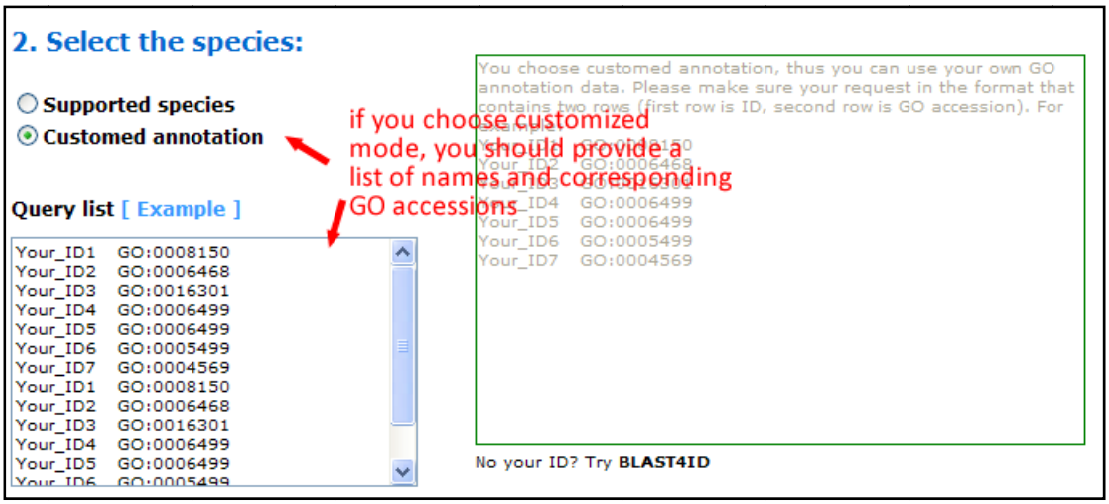
SEA is a traditional and widely used method. It is simple to use and simple to understand. User only needs to prepare a list of gene/probe names, and enrichment GO terms will be found out after statistical test from pre-calculated background or customized one.

STEP 1:

To use SEA analysis, you should firstly select the type of your query list, either single names or names with GO accession. If you choose using supported species in agriGO, you only need to provide a list of sequence identifiers. It should be noted that you would better select species and check all allowed ID types of corresponding species, then submit your IDs. Only allowed IDs are suitable to be analysis in this type mode. And you can mix your IDs from different types. Just ensure they are allowed IDs.



If you choose customized mode, you are no long limited by agriGO-owned species any more. You can use any IDs you have, but only be noticed IDs should attach with GO accession!

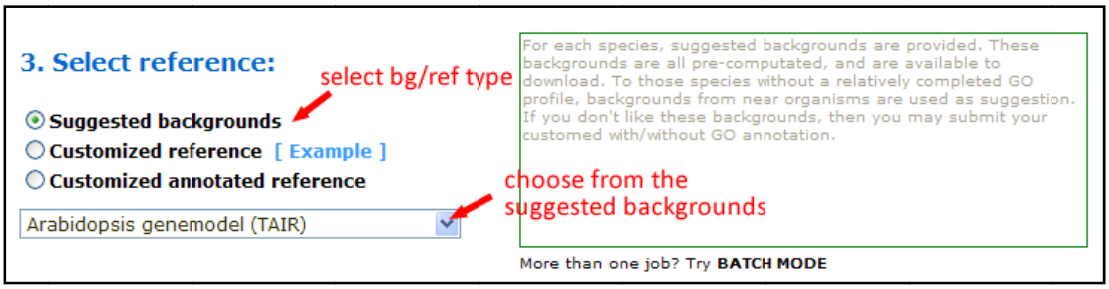

2. Select the species:
Supported species
Customed annotation
if you choose customized mode, you should provide a list of names and corresponding GO accessions
Query list [Example]
Your_ID1 GO:0008150
Your_ID2 GO:0006468
Your_ID3 GO:0016301
Your_ID4 GO:0006499
Your_ID5 GO:0006499
Your_ID6 GO:0005499
Your_ID7 GO:0004569
Your_ID1 GO:0008150
Your_ID2 GO:0006468
Your_ID3 GO:0016301
Your_ID4 GO:0006499
Your_ID5 GO:0006499
You choose customed annotation, thus you can use your own GO annotation data. Please make sure your request in the format that contains two rows (first row is ID, second row is GO accession). For
Your_ID4 GO:0006499
Your_ID5 GO:0006499
Your_ID6 GO:0005499
Your_ID7 GO:0004569
No your ID? Try BLAST4ID

OK, theoretically you can just click “Submit” to perform analysis now and simply skip following steps. Nevertheless, if you want set more advanced parameters, then keep on reading this manual.

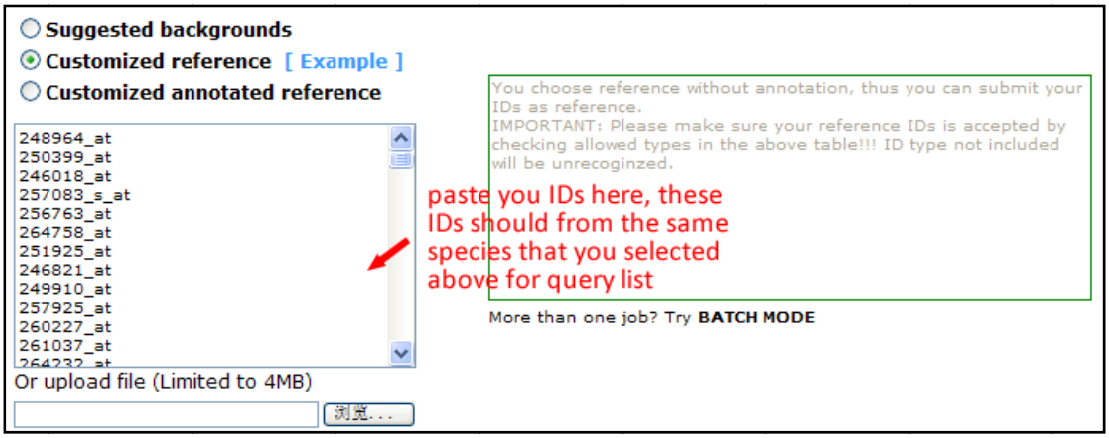
STEP 2:

Now you can set the background or reference. There three types: suggested backgrounds, customized reference and customized annotated reference.

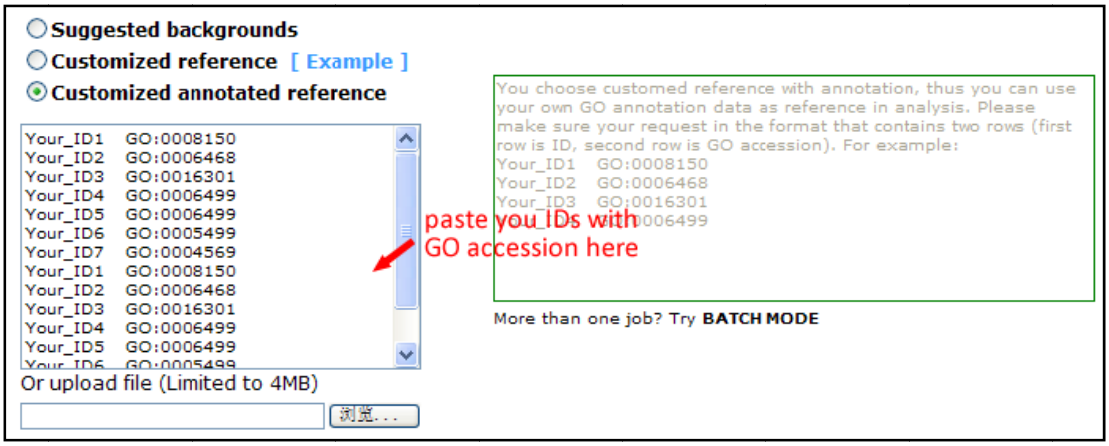
The default parameter is using suggested backgrounds. For each species, agriGO will give all possible the background types. To those species without a relatively completed profile, backgrounds from neighbored organisms are suggested. Users can select based on their practical need, otherwise use customized reference.


3. Select reference:
select bg/ref type
Suggested backgrounds
Customized reference [Example]
Customized annotated reference
Arabidopsis genemodel (TAIR)
choose from the suggested backgrounds
For each species, suggested backgrounds are provided. These backgrounds are all pre-computated, and are available to download. To those species without a relatively completed GO profile, backgrounds from near organisms are used as suggestion. If you don't like these backgrounds, then you may submit your customed with/without GO annotation.
More than one job? Try BATCH MODE

In the case that you do not want any of suggested pre-computated background, you can use customized reference instead. NOTE: IDs in reference list should from the same species that one selected above for query list.


Suggested backgrounds
Customized reference [Example]
Customized annotated reference
248964_at
250399_at
246018_at
257083_s_at
256763_at
264758_at
251925_at
246821_at
249910_at
257925_at
260227_at
261037_at
Or upload file (Limited to 4MB)
浏览...
paste you IDs here, these IDs should from the same species that you selected above for query list
You choose reference without annotation, thus you can submit your IDs as reference.
IMPORTANT: Please make sure your reference IDs is accepted by checking allowed types in the above table!!! ID type not included will be unrecoginzed.
More than one job? Try BATCH MODE

Also you can use any IDs if you choose customized annotated reference mode, however, the price is to attach with GO accession to obtain such freedom. ☺


Suggested backgrounds
Customized reference [Example]
Customized annotated reference
Your_ID1 GO:0008150
Your_ID2 GO:0006468
Your_ID3 GO:0016301
Your_ID4 GO:0006499
Your_ID5 GO:0006499
Your_ID6 GO:0005499
Your_ID7 GO:0004569
Your_ID1 GO:0008150
Your_ID2 GO:0006468
Your_ID3 GO:0016301
Your_ID4 GO:0006499
Your_ID5 GO:0006499
Or upload file (Limited to 4MB)
浏览...
paste you IDs with GO accession here
You choose customed reference with annotation, thus you can use your own GO annotation data as reference in analysis. Please make sure your request in the format that contains two rows (first row is ID, second row is GO accession). For example:
Your_ID1 GO:0008150
Your_ID2 GO:0006468
Your_ID3 GO:0016301
More than one job? Try BATCH MODE

You can paste direct or upload your file, for latter, please make sure the file no bigger than 4MB.

STEP 3:

The advanced options are optional but quite important. These options are default hidden, and need to one click to make them visible. In SEA analysis, there are three statistical test methods: hypergeometric, chi-square and fisher test.

When the input/query list is compared with the previously computed background, or is a subset of reference list, choose hypergeometric or fisher. When both of your query list number and reference list number are quite small, you may better choose fisher test. When the input/query list has few or no intersections with the reference list, the Chi-square tests are more appropriate.

Next you can choose method to do the multi-test adjustment. Seven adjustment methods are available here, including: Yekutieli (FDR under dependency), Bonferroni, Hochberg, Hochberg (FDR), Hommel, Holm, False Discovery Rate. Though I would suggest perform adjustment test, you truly can turn off it and use no adjust. While you choose no adjust, then you may set significant level below higher. Terms under the cutoff of the significant level will be highlighted, and emphasized in analysis results, and it will affect your test output.

Minimum number of mapping entries means that GO annotations that do not appear in at least the selected number of entries will not be shown. In other word, higher you set the number, more entries needed to make one GO term appear in the analysis result.

Gene ontology type: Plant GO slim is a cut-down version of the GO ontologies containing a subset of the terms in the whole GO for plant.

Last, if you provide a mail address, a notification will be send when the analysis is completed with the link to the results. Providing a email address is optional to SEA analysis, because it is very fast.

4. Advanced options (optional):
Statistical test method
Hypergeometric
Multi-test adjustment method
Yekutieli (FDR under dependency)
Significance Level
0.05
Minimum number of mapping entries
5
Gene ontology type
Complete GO
Plant GO slim
Email notification
Statistical test: When the input list is compared with the previously computed background, or is a subset of reference list, choose hypergeometric or fisher, for latter only when your query number is quite small. When the input list has few or no intersections with the reference list, the Chi-square tests are more appropriate.
Multi-test adjustment method: Choose method to do the multi-test adjustment. If you choose no adjust, then you may set significant level higher.
Significance Level: Terms under the cutoff will be highlighted, and emphasized in analysis results.
Minimum number of mapping entries: GO Annotations that do not appear in at least the selected number of entries will not be shown.
Gene ontology type: Plant GO slim is a cut-down version of the GO ontologies containing a subset of the terms in the whole GO for plant.
Email: If you provide a mail address a notification will be send when the analysis is completed with the link to the results.

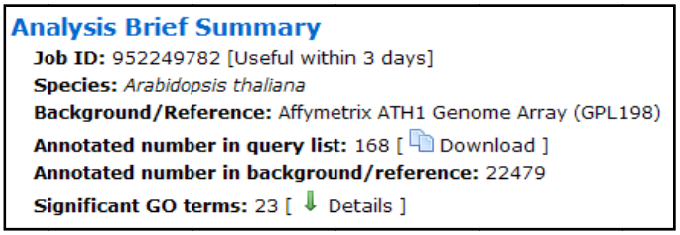
Greeting! You can now click submit to perform the analysis. You can always get interactive help from the right help frames, and a detailed tutorial in this manual, if you still have any question then contact me directly.

In the following we will discuss the outputs of the SEA analysis.

Singular Enrichment Analysis (SEA) Results

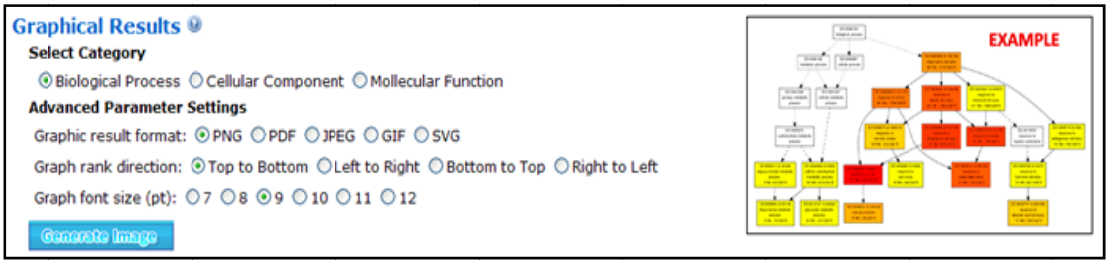
Part 1:

A brief summary of your job will be given. The job ID is useful within 3 days. A file containing all entities in the query list that can be annotated by GO associated with descriptions is able to download.


Analysis Brief Summary
Job ID: 952249782 [Useful within 3 days]
Species: Arabidopsis thaliana
Background/Reference: Affymetrix ATH1 Genome Array (GPL198)
Annotated number in query list: 168 [Download]
Annotated number in background/reference: 22479
Significant GO terms: 23 [Details]

Part 2:

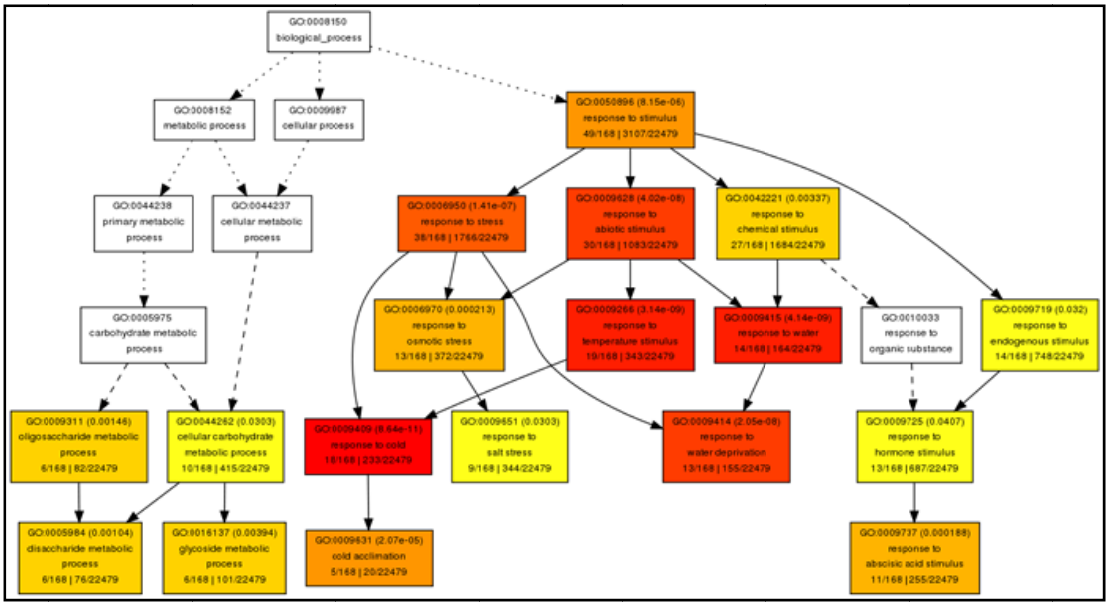
In this part, you can browse the hieratical graph result. Note that the graphical result was generated as separate graphs for each of the three GO categories, namely biological process, molecular function and cellular component. After select the category, uses can specified their favorite output format, graph rank direction and font size. The result format means which output format you preferred. The rank direction is used to define the direction in your output, for instance the direction in the example image is 'top to bottom'. And the font size is self-evident that user can set smaller size if there are many nodes in their result.


Graphical Results
Select Category
Biological Process
Cellular Component
Mollecular Function
Advanced Parameter Settings
Graphic result format: PNG PDF JPEG GIF SVG
Graph rank direction: Top to Bottom Left to Right Bottom to Top Right to Left
Graph font size (pt): 7 8 9 10 11 12
Generate Image
EXAMPLE

Click the 'generate image' bottom after you set all parameters well. The graphical result will be presented according to your own settings. The graphical result is a GO hieratical image containing all statistically significant terms.

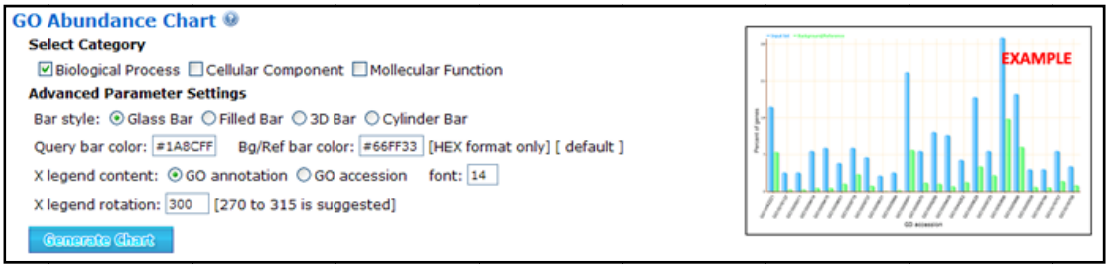
These nodes in the image are classified into ten levels which are associated with corresponding specific colors. The smaller of the term's q-value, the more significant statistically, and the node's color is darker and redder (Note: q-value here means that the value of the multiple-test adjusted p-value). Inside the box of the significant terms, the information includes: GO term, q-value, GO description, item number mapping the GO in the query list and background, and total number of query list and background. But when those term whose q-value is higher than the cutoff set by the user, only GO information will be given in the box.

To better understand the graphical result, investigation of the annotation diagram is suggested. If user chooses PNG or JPG or GIF result format, linkage to the term's detail is available by clicking those blocks.



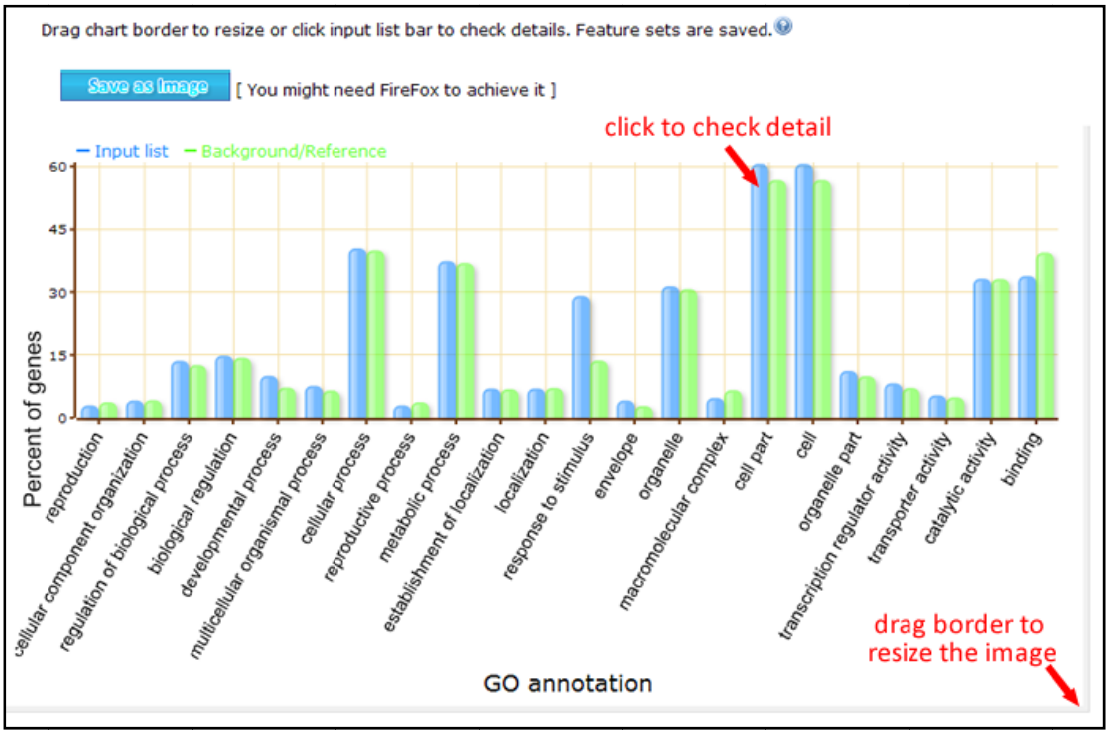
Part 3:

The terms selected here are children terms of root one (or called secondary level terms) or significant terms of secondary level terms. Thus, the bar chart gives user a brief portray since the GO terms are relatively general description. Similar to the procedure of graphical result, user should specified their parameters before create the GO abundance chart. User can try these setting to obtain favorite view of the chart bar. Note the setting you used will be recorded in your cookie and these settings will be default ones in your future jobs. In other word, you may try several times and make your last attempt as your own features.


GO Abundance Chart
Select Category
Biological Process
Cellular Component
Mollecular Function
Advanced Parameter Settings
Bar style: Glass Bar Filled Bar 3D Bar Cylinder Bar
Query bar color: #1A8CFF
Bg/Ref bar color: #66FF33 [HEX format only] [default]
X legend content: GO annotation GO accession font: 14
X legend rotation: 300 [270 to 315 is suggested]
Generate Chart
EXAMPLE

Here the bar chart is using glass bar style, default colors, GO annotation as X legend, 14px font and 300 for X legend rotation. Here are some tips: 1. Have a glance of all four bar styles and select one you like, 2. Use HEX format to define colors and there is a website we already suggested, 3. If you prefer GO annotation as X legend content, you may use smaller font once there are too many words, 4. 270 to 315 is suggested for X legend rotation, in which 270 means vertical, and 315 means 45 degree slope, and you can try other number which may satisfy your taste but seems somehow strange to me ☺.

The bar chart is created based on scripts from Open Flash Chart. It is powerful. You can drag borders to resize and adjust the image size and ratio. And bars are accessible to term's detailed information. A 'Save as Image' bottom is existed but only useful when you are using FireFox browser, and if you can also use your 'Print Screen' bottom on your keyboard or other tools to download this image.

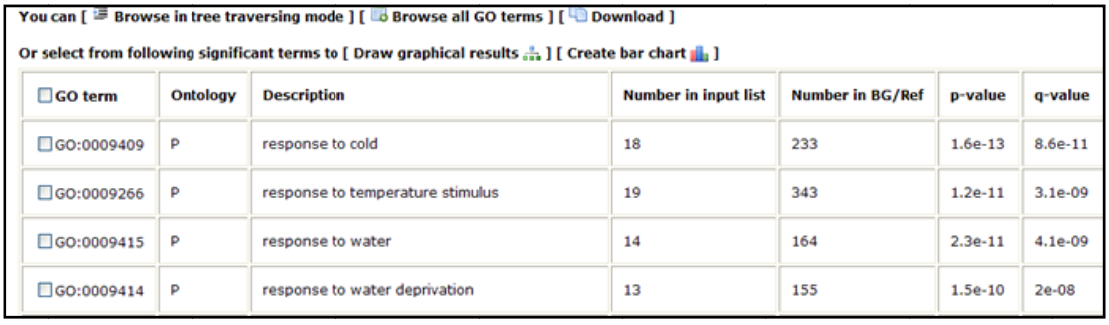

Drag chart border to resize or click input list bar to check details. Feature sets are saved.
Save as Image
[You might need FireFox to achieve it]
click to check detail
Input list
Background/Reference
Percent of genes
GO annotation
drag border to resize the image

Part 4:

In this part, detailed information is given. All GO significant terms will presented in the following table. And you can browse the GO terms using tree traversing mode (we will discussed it later), or can browse all GO terms in the similar type table, or just the data.

User can select terms to draw graphical result or create bar chart. Please note that the parameters used in graph or chart generating is fetched from your cookie, and your cookie will be set or changed when you generate graphical results or GO abundant chart which has been mentioned in part 2 and 3. While it will make you a bit trouble if you would like adjust the images created here to redo the part 2 or 3 work once more to change the settings.

Click the checkbox left to 'GO term' can select all GO terms at one time.

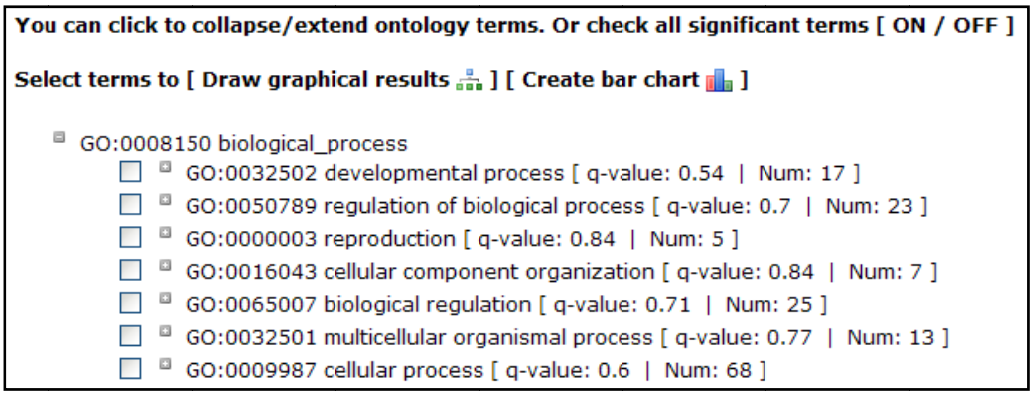


You can [Browse in tree traversing mode] [Browse all GO terms] [Download]

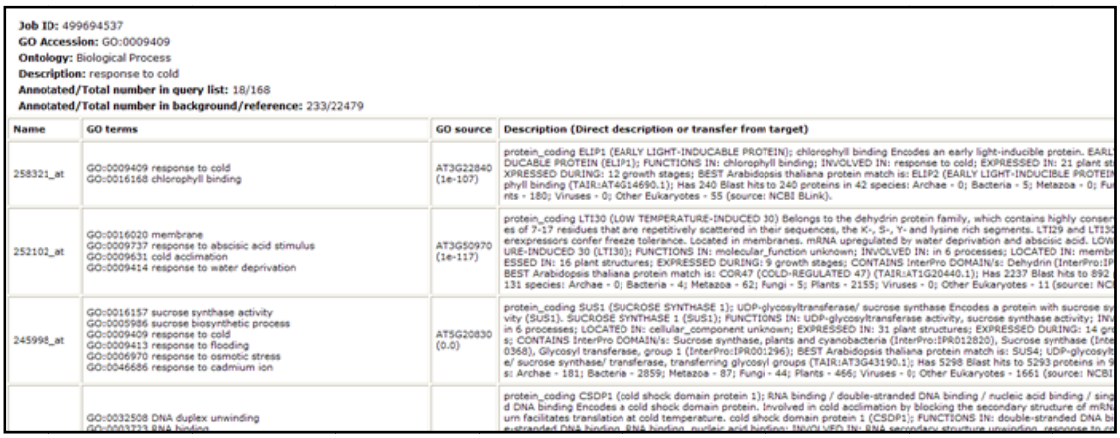
Or select from following significant terms to [Draw graphical results] [Create bar chart]

GO term	Ontology	Description	Number in input list	Number in BG/Ref	p-value	q-value
GO:0009409	P	response to cold	18	233	1.6e-13	8.6e-11
GO:0009266	P	response to temperature stimulus	19	343	1.2e-11	3.1e-09
GO:0009415	P	response to water	14	164	2.3e-11	4.1e-09
GO:0009414	P	response to water deprivation	13	155	1.5e-10	2e-08

You can click the GO name to collapse/extend ontology terms in tree traversing mode. A bottom that can make all significant selected or not is available and those selected terms can be used in drawing graphical results or to create bar chart. Please note that at least one significant term should be included in graph generation, otherwise the graphical result will be some kind blank and meaningless. Click on the number will lead you to term's detail information.


You can click to collapse/extend ontology terms. Or check all significant terms [ON / OFF]
Select terms to [Draw graphical results] [Create bar chart]
GO:0008150 biological_process
GO:0032502 developmental process [q-value: 0.54 | Num: 17]
GO:0050789 regulation of biological process [q-value: 0.7 | Num: 23]
GO:0000003 reproduction [q-value: 0.84 | Num: 5]
GO:0016043 cellular component organization [q-value: 0.84 | Num: 7]
GO:0065007 biological regulation [q-value: 0.71 | Num: 25]
GO:0032501 multicellular organismal process [q-value: 0.77 | Num: 13]
GO:0009987 cellular process [q-value: 0.6 | Num: 68]

The term's detail page is as following. The agriGO will give all entries can be annotated to the term besides a brief summary. And for each entry the annotation includes: GO terms, GO source, description.



Job ID: 499694537
GO Accession: GO:0009409
Ontology: Biological Process
Description: response to cold
Annotated/Total number in query list: 18/168
Annotated/Total number in background/reference: 233/22479

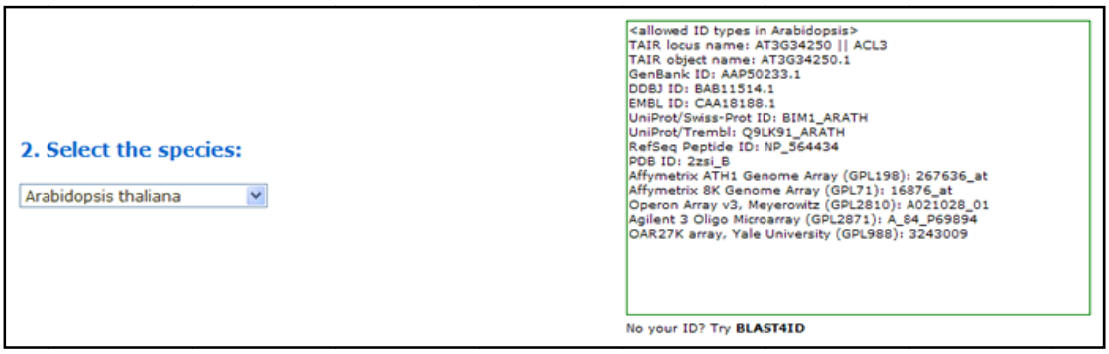
Name	GO terms	GO source	Description (Direct description or transfer from target)
258321_at	GO:0009409 response to cold GO:0016168 chlorophyll binding	AT3G22840 (1e-107)	protein_coding ELIP1 (EARLY LIGHT-INDUCABLE PROTEIN); chlorophyll binding Encodes an early light-inducible protein. EARL DUCABLE PROTEIN (ELIP1); FUNCTIONS IN: chlorophyll binding; INVOLVED IN: response to cold; EXPRESSED IN: 21 plant st XPRESSED DURING: 12 growth stages; BEST Arabidopsis thaliana protein match is: ELIP2 (EARLY LIGHT-INDUCIBLE PROTEIN phyll binding (TAIR:AT4G14690.1); Has 240 Blast hits to 240 proteins in 42 species: Archae - 0; Bacteria - 5; Metazoa - 0; Fu nts - 180; Viruses - 0; Other Eukaryotes - 55 (source: NCBI BLink).
252102_at	GO:0016020 membrane GO:0009737 response to abscisic acid stimulus GO:0009631 cold acclimation GO:0009414 response to water deprivation	AT3G50970 (1e-117)	protein_coding LTI30 (LOW TEMPERATURE-INDUCED 30) Belongs to the dehydrin protein family, which contains highly conser es of 7-17 residues that are repetitively scattered in their sequences, the K-, S-, Y- and lysine rich segments. LTI29 and LTI30 erexpressors confer freeze tolerance. Located in membranes. mRNA upregulated by water deprivation and abscisic acid. LOW URE-INDUCED 30 (LTI30); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: in 6 processes; LOCATED IN: membr ESSED IN: 16 plant structures; EXPRESSED DURING: 9 growth stages; CONTAINS InterPro DOMAIN/s: Dehydrin (InterPro:IP BEST Arabidopsis thaliana protein match is: COR47 (COLD-REGULATED 47) (TAIR:AT1G20440.1); Has 2237 Blast hits to 892 131 species: Archae - 0; Bacteria - 4; Metazoa - 62; Fungi - 5; Plants - 2155; Viruses - 0; Other Eukaryotes - 11 (source: NC
245998_at	GO:0016157 sucrose synthase activity GO:0005986 sucrose biosynthetic process GO:0009409 response to cold GO:0009413 response to flooding GO:0006970 response to osmotic stress GO:0046686 response to cadmium ion	AT5G20830 (0.0)	protein_coding SUS1 (SUCROSE SYNTHASE 1); UDP-glycosyltransferase/ sucrose synthase Encodes a protein with sucrose sy vity (SUS1). SUCROSE SYNTHASE 1 (SUS1); FUNCTIONS IN: UDP-glycosyltransferase activity, sucrose synthase activity; INV in 6 processes; LOCATED IN: cellular_component unknown; EXPRESSED IN: 31 plant structures; EXPRESSED DURING: 14 gr s; CONTAINS InterPro DOMAIN/s: Sucrose synthase, plants and cyanobacteria (InterPro:IPR012820), Sucrose synthase (Inte 0368), Glycosyl transferase, group 1 (InterPro:IPR001296); BEST Arabidopsis thaliana protein match is: SUS4; UDP-glycosylt e/ sucrose synthase/ transferase, transferring glycosyl groups (TAIR:AT3G43190.1); Has 5298 Blast hits to 5293 proteins in 9 s: Archae - 181; Bacteria - 2859; Metazoa - 87; Fungi - 44; Plants - 466; Viruses - 0; Other Eukaryotes - 1661 (source: NCBI
	GO:0032508 DNA duplex unwinding GO:0003723 RNA binding		protein_coding CSDP1 (cold shock domain protein 1); RNA binding / double-stranded DNA binding / nucleic acid binding / sing d DNA binding Encodes a cold shock domain protein. Involved in cold acclimation by blocking the secondary structure of mRN urn facilitates translation at cold temperature. cold shock domain protein 1 (CSDP1); FUNCTIONS IN: double-stranded DNA bi e-stranded DNA binding, RNA binding, nucleic acid binding; INVOLVED IN: RNA secondary structure unwinding, response to co

B. How to use Parametric Analysis of Gene Set Enrichment (PAGE)?

PAGE method is argued by Kim [*BMC Bioinformatics* 2005, 6:144]. Using Central Limit Theorem in statistics, this method is simple and efficient. Different to SEA, it takes expression level into account, and can deal with a long list of genes/probesets.

STEP 1:

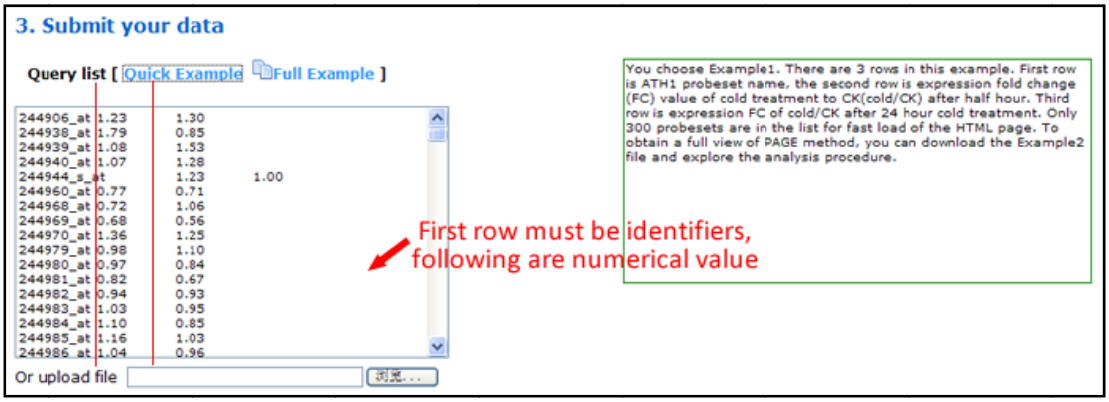
Firstly, you should choose the species for your query data. Please make sure that identifiers in your input should be one of datatypes inside the right information table. If your identifiers are not stored in agriGO, there is another two ways: one is provided your own GO annotation file, the other is to use our BLAST4ID service.


2. Select the species:
Arabidopsis thaliana
<allowed ID types in Arabidopsis>
TAIR locus name: AT3G34250 || ACL3
TAIR object name: AT3G34250.1
GenBank ID: AAP50233.1
DDBJ ID: BAB11514.1
EMBL ID: CAA18188.1
UniProt/Swiss-Prot ID: BIM1_ARATH
UniProt/Trembl: Q9LK91_ARATH
RefSeq Peptide ID: NP_564434
PDB ID: 2zsi_B
Affymetrix ATH1 Genome Array (GPL198): 267636_at
Affymetrix 8K Genome Array (GPL71): 16876_at
Operon Array v3, Meyerowitz (GPL2810): A021028_01
Agilent 3 Oligo Microarray (GPL2871): A_84_P69894
OAR27K array, Yale University (GPL988): 3243009
No your ID? Try BLAST4ID

STEP 2:

In PAGE analysis, user should pay more attention to input data. As presented in the following image, as least two rows must be provided. The first row is sequence identifiers, and followings are numerical value. The numerical value is fold change (FC) or log2-transformed FC value (latter preferred) of the identifiers' expression under different condition. If you do not have expression data, then SEA may be the alternative choice.

In agriGO's example, there are 3 rows in this example. First row is ATH1 probeset name, the second row is expression fold change (FC) value of cold treatment to CK(cold/CK) after half hour. Third row is expression FC of cold/CK after 24 hour cold treatment. Only 600 probesets are in the quick example for the fast load of the HTML page. To obtain a full view of PAGE method, you can download the full example file and explore the following analysis procedure.


3. Submit your data
Query list [Quick Example Full Example]
244906_at 1.23 1.30
244938_at 1.79 0.85
244939_at 1.08 1.53
244940_at 1.07 1.28
244944_s_at 1.23 1.00
244960_at 0.77 0.71
244968_at 0.72 1.06
244969_at 0.68 0.56
244970_at 1.36 1.25
244979_at 0.98 1.10
244980_at 0.97 0.84
244981_at 0.82 0.67
244982_at 0.94 0.93
244983_at 1.03 0.95
244984_at 1.10 0.85
244985_at 1.16 1.03
244986_at 1.04 0.96
Or upload file
First row must be identifiers, following are numerical value
You choose Example1. There are 3 rows in this example. First row is ATH1 probeset name, the second row is expression fold change (FC) value of cold treatment to CK(cold/CK) after half hour. Third row is expression FC of cold/CK after 24 hour cold treatment. Only 300 probesets are in the list for fast load of the HTML page. To obtain a full view of PAGE method, you can download the Example2 file and explore the analysis procedure.

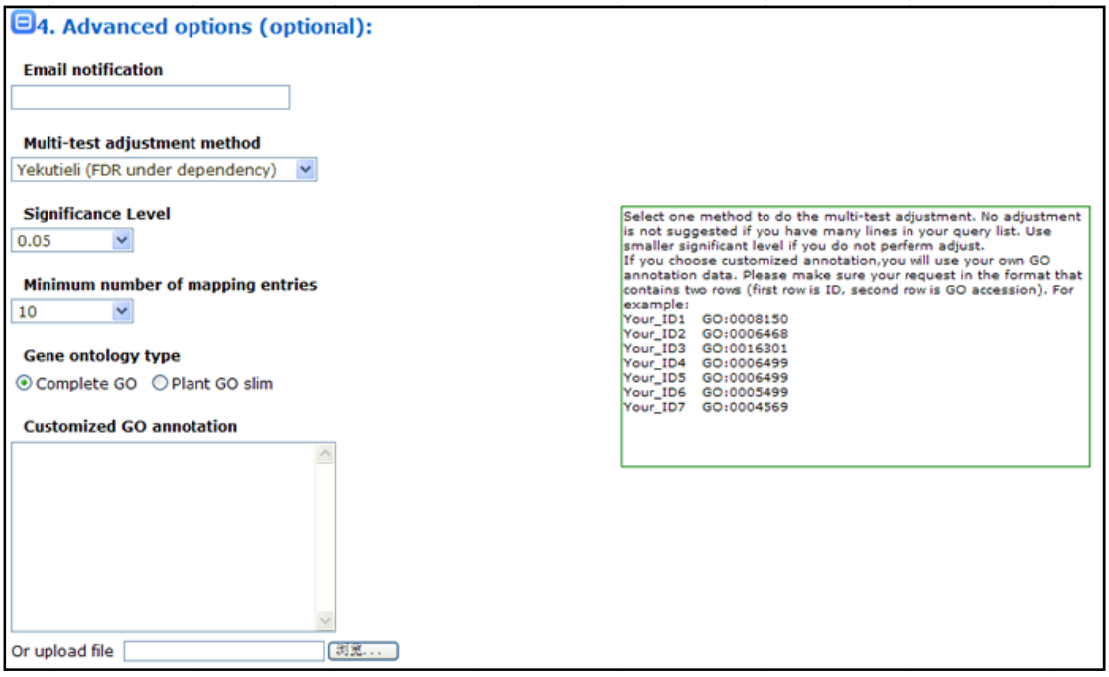
STEP 3:

Next you can choose method to do the multi-test adjustment. Seven adjustment methods are available here, including: Yekutieli (FDR under dependency), Bonferroni, Hochberg, Hochberg (FDR), Hommel, Holm, False Discovery Rate. Though I would suggest perform adjustment test, you truly can turn off it and use no adjust. While you choose no adjust, then you may set significant level below higher. Terms under the cutoff of the significant level will be highlighted, and emphasized in analysis results, and it will affect your test output.

Minimum number of mapping entries means that GO annotations that do not appear in at least the selected number of entries will not be shown. In other word, higher you set the number, more entries needed to make one GO term appear in the analysis result.

Gene ontology type: Plant GO slim is a cut-down version of the GO ontologies containing a subset of the terms in the whole GO for plant.

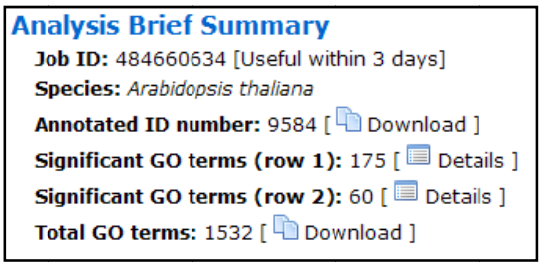
If you can also upload your own customized GO annotation file once your identifiers are not accepted directly by agriGO. The file's size is limited to 4MB.


4. Advanced options (optional):
Email notification
Multi-test adjustment method
Yekutieli (FDR under dependency)
Significance Level
0.05
Minimum number of mapping entries
10
Gene ontology type
Complete GO
Plant GO slim
Customized GO annotation
Or upload file
浏览...
Select one method to do the multi-test adjustment. No adjustment is not suggested if you have many lines in your query list. Use smaller significant level if you do not perferm adjust.
If you choose customized annotation,you will use your own GO annotation data. Please make sure your request in the format that contains two rows (first row is ID, second row is GO accession). For example:
Your_ID1 GO:0008150
Your_ID2 GO:0006468
Your_ID3 GO:0016301
Your_ID4 GO:0006499
Your_ID5 GO:0006499
Your_ID6 GO:0005499
Your_ID7 GO:0004569

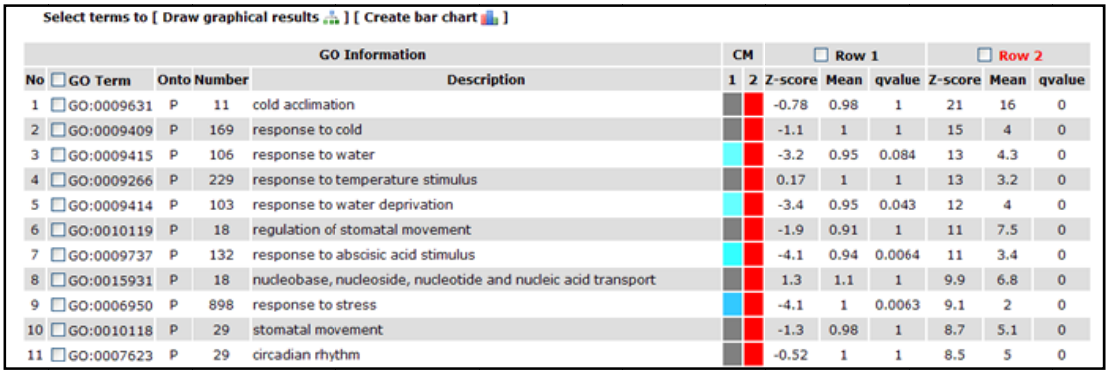
OK, now you can click submit to start analysis now. You may explore output of analysis results in the following part of manual.

Parametric Analysis of Gene Set Enrichment (PAGE) Result

Results generated by PAGE analysis have many similar points to SEA analysis, thus it is suggested to browse SEA result introduction part firstly. And only unique features to PAGE results will be explained. Since PAGE tool can analysis several rows at one time, and terms in each row will be calculated, each row has its significant GO terms. Number of significant GO terms for each row is listed in the brief summary part.


Analysis Brief Summary
Job ID: 484660534 [Useful within 3 days]
Species: Arabidopsis thaliana
Annotated ID number: 9584 [Download]
Significant GO terms (row 1): 175 [Details]
Significant GO terms (row 2): 60 [Details]
Total GO terms: 1532 [Download]

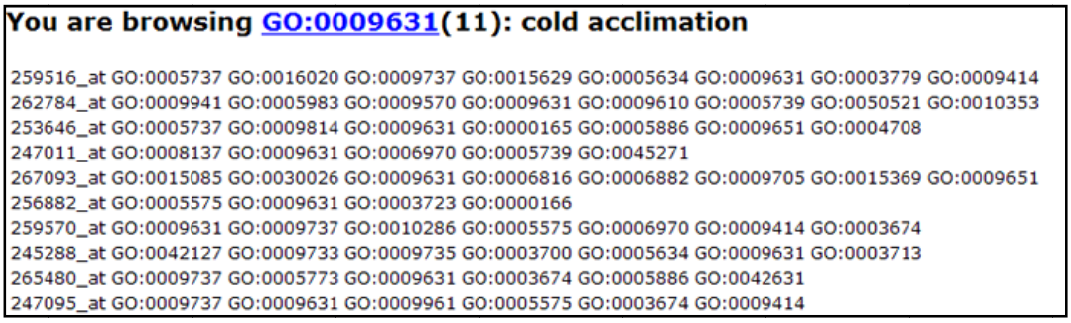
The number of terms is determined by the row you selected which is colored by red. A simple colorful model named CM for short is available. The color used in the CM is same to the color used in graphical result in which red color system means up regulated and blue means down regulated. And each block present the term's Z-score for the row. You can select row(s) and term(s) to generate further images.



Select terms to [Draw graphical results] [Create bar chart]

	GO Information				CM		Row 1			Row 2		
No	GO Term	Onto	Number	Description	1	2	Z-score	Mean	qvalue	Z-score	Mean	qvalue
1	GO:0009631	P	11	cold acclimation			-0.78	0.98	1	21	16	0
2	GO:0009409	P	169	response to cold			-1.1	1	1	15	4	0
3	GO:0009415	P	106	response to water			-3.2	0.95	0.084	13	4.3	0
4	GO:0009266	P	229	response to temperature stimulus			0.17	1	1	13	3.2	0
5	GO:0009414	P	103	response to water deprivation			-3.4	0.95	0.043	12	4	0
6	GO:0010119	P	18	regulation of stomatal movement			-1.9	0.91	1	11	7.5	0
7	GO:0009737	P	132	response to abscisic acid stimulus			-4.1	0.94	0.0064	11	3.4	0
8	GO:0015931	P	18	nucleobase, nucleoside, nucleotide and nucleic acid transport			1.3	1.1	1	9.9	6.8	0
9	GO:0006950	P	898	response to stress			-4.1	1	0.0063	9.1	2	0
10	GO:0010118	P	29	stomatal movement			-1.3	0.98	1	8.7	5.1	0
11	GO:0007623	P	29	circadian rhythm			-0.52	1	1	8.5	5	0

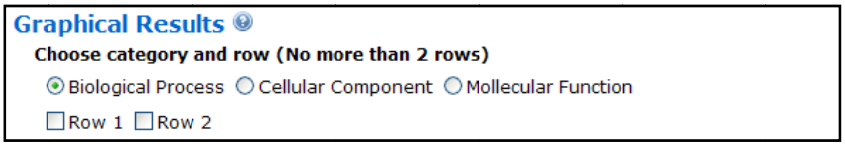
The term's detailed information is generated if you click the number. This page may be a bit simple because it is quite possible that there are too many entries mapping to the GO.



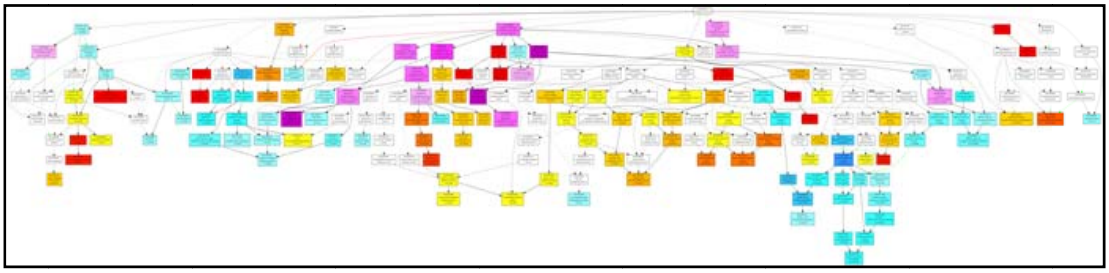
You are browsing GO:0009631(11): cold acclimation

259516_at GO:0005737 GO:0016020 GO:0009737 GO:0015629 GO:0005634 GO:0009631 GO:0003779 GO:0009414
262784_at GO:0009941 GO:0005983 GO:0009570 GO:0009631 GO:0009610 GO:0005739 GO:0050521 GO:0010353
253646_at GO:0005737 GO:0009814 GO:0009631 GO:0000165 GO:0005886 GO:0009651 GO:0004708
247011_at GO:0008137 GO:0009631 GO:0006970 GO:0005739 GO:0045271
267093_at GO:0015085 GO:0030026 GO:0009631 GO:0006816 GO:0006882 GO:0009705 GO:0015369 GO:0009651
256882_at GO:0005575 GO:0009631 GO:0003723 GO:0000166
259570_at GO:0009631 GO:0009737 GO:0010286 GO:0005575 GO:0006970 GO:0009414 GO:0003674
245288_at GO:0042127 GO:0009733 GO:0009735 GO:0003700 GO:0005634 GO:0009631 GO:0003713
265480_at GO:0009737 GO:0005773 GO:0009631 GO:0003674 GO:0005886 GO:0042631
247095_at GO:0009737 GO:0009631 GO:0009961 GO:0005575 GO:0003674 GO:0009414

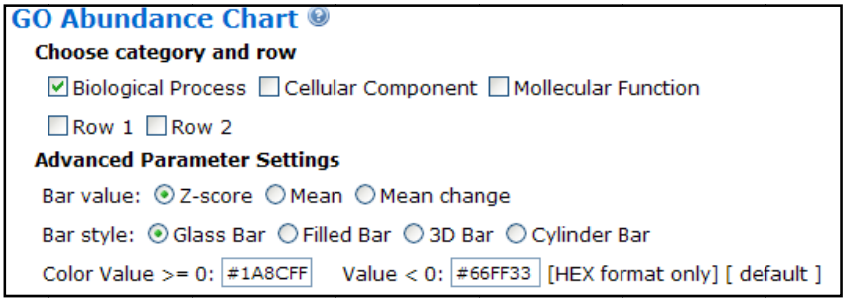
In graphical result part, user can choose one or two rows to draw the image. If two rows are selected, a third color system (purple colors) will be used in demonstrating those terms have different regulation direction in two rows.



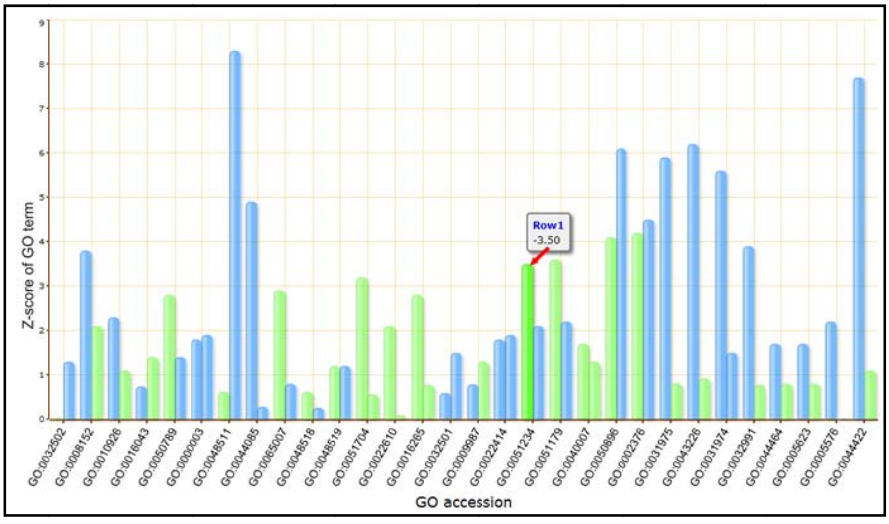
The following example presents two rows in one graph. You can check the annotation diagram below the result. There are three color systems: red means up regulated terms, blue means down regulated and purple presents the term is regulated in different direction in tow rows. And if the term has same regulated direction in both rows, it will have double borders. In the box 'r1=1e-10' means the q-value of the term in row1 is 1e-10, and 'zs' presents Z-score.



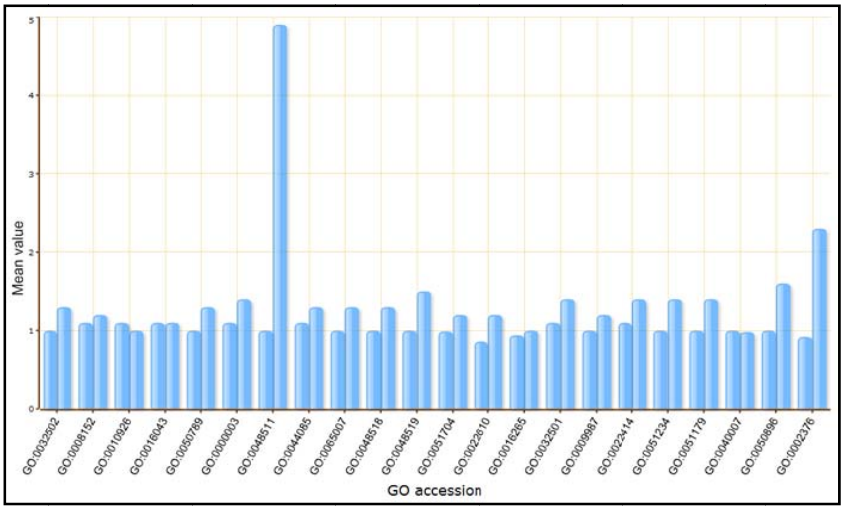
In bar chart generation part, Z-score is the statistical value in PAGE calculation, mean value is the mean of the value of all entries in the row. Mean change is mean minus standard deviation which presents the change of expression when comparing to the whole row background. While user can set two color values for up-regulation terms and down-regulation terms.



As mentioned before, Z-score which is bigger than 0 or smaller than 0 will be presented using different colors which set by user.



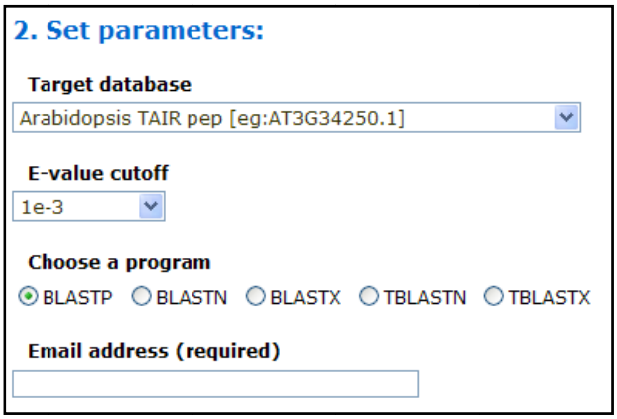
But if you choose mean value, they are in the same color since all mean is bigger than 0.



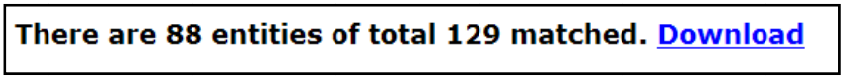
C. How to use BLAST4ID tool?

The BLAST4ID tool is not an analysis tool, but an associated one used mainly for two purposes: 1. Transfer your IDs which are not available to agriGO to available ones, 2. use blast search to annotate your sequences with GO.

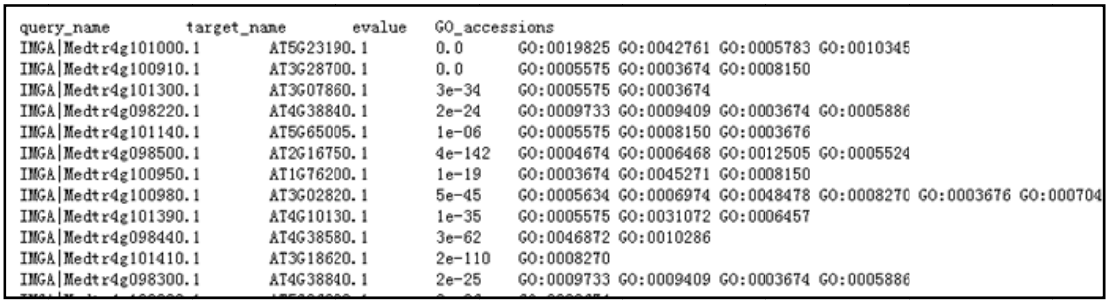
To use BLAST4ID, user should set target database at first, and then E-value cutoff. The program should be correctly selected based on sequence types of user's input and target database. Generally speaking, all array sequences are nucleotide and other genome sequences are protein. The process may take a long while thus the E-mail address should be given for the Email notification.



The result interface is a bit simple☺, but enough for usage.



Downloadable text result:



query_name	target_name	evalue	GO_accessions
IMGA|Medtr4g101000.1	AT5G23190.1	0.0	GO:0019825 GO:0042761 GO:0005783 GO:0010345
IMGA|Medtr4g100910.1	AT3G28700.1	0.0	GO:0005575 GO:0003674 GO:0008150
IMGA|Medtr4g101300.1	AT3G07860.1	3e-34	GO:0005575 GO:0003674
IMGA|Medtr4g098220.1	AT4G38840.1	2e-24	GO:0009733 GO:0009409 GO:0003674 GO:0005886
IMGA|Medtr4g101140.1	AT5G65005.1	1e-06	GO:0005575 GO:0008150 GO:0003676
IMGA|Medtr4g098500.1	AT2G16750.1	4e-142	GO:0004674 GO:0006468 GO:0012505 GO:0005524
IMGA|Medtr4g100950.1	AT1G76200.1	1e-19	GO:0003674 GO:0045271 GO:0008150
IMGA|Medtr4g100980.1	AT3G02820.1	5e-45	GO:0005634 GO:0006974 GO:0048478 GO:0008270 GO:0003676 GO:000704
IMGA|Medtr4g101390.1	AT4G10130.1	1e-35	GO:0005575 GO:0031072 GO:0006457
IMGA|Medtr4g098440.1	AT4G38580.1	3e-62	GO:0046872 GO:0010286
IMGA|Medtr4g101410.1	AT3G18620.1	2e-110	GO:0008270
IMGA|Medtr4g098300.1	AT4G38840.1	2e-25	GO:0009733 GO:0009409 GO:0003674 GO:0005886

D. How to use search tool?

Search tools in agriGO are easy to understand and use. Here are some tips:

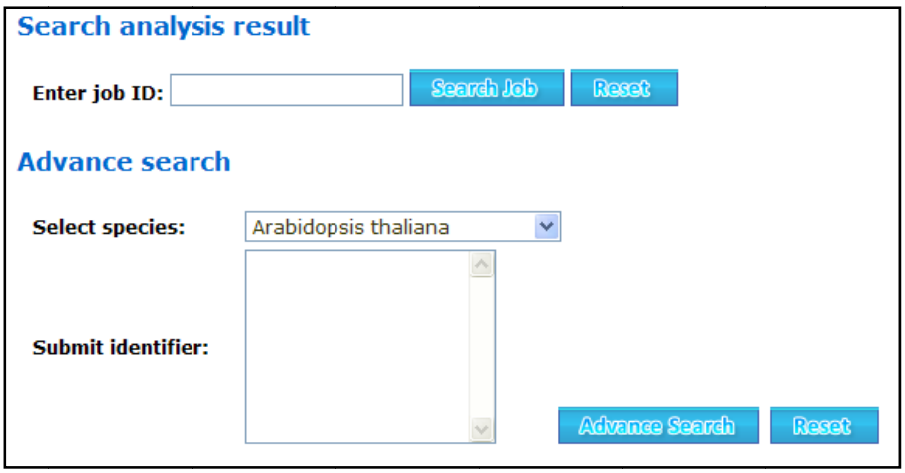
1. Unless you contact me and ask for elongation, the job ID is available within 3 days.

2. There is a short-cut at top-right corner for job searching.

3. In advance search, you have to define the species firstly.

4. You can either search single one or a list (no more than 100) of sequence identifiers.

5. Input IDs are case insensitive, but will be agriGO's format in the output.



6. FAQ

What is agriGO?

The agriGO is designed to automate the job for experimental biologists to identify enriched Gene Ontology (GO) terms in a list of microarray probe sets or gene identifiers (with or without expression information) and it is also a GO-related database. The agriGO specially focus on agricultural species.

What is GO?

"The Gene Ontology (GO) project provides a controlled vocabulary to describe gene and gene product attributes in any organism. The GO project is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The GO collaborators are developing three structured, controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. There are three separate aspects to this effort: first, we write and maintain the ontologies themselves; second, we make cross-links between the ontologies and the genes and gene products in the collaborating databases, and third, we develop tools that facilitate the creation, maintainence and use of ontologies." Definition from http://www.geneontology.org/

What is updated in agriGO compare with EasyGO?

The agriGO is a successor of EasyGO, and it go further. 1. We create new website interface. The database structure and scripts of agriGO are redesigned. Both page loading speed and analysis speed of agriGO now are improved because of the change. 2. The agriGO service is especially focus on agricultural species. It supports species is extended to 35 including as much as 280 datatypes. 3. We added new analysis tools for new agriGO, such as PAGE analysis and BLAST4ID tool. 4. The result output and information are richer compare with EasyGO. 5. The agriGO could also work as a GO database with search and download service.

What are the unique features of agriGO compare with other GO webserver/database?

The agriGO provides heavy support to agricultural species. Not only limited to SEA analysis, GSEA which is achieved using PAGE method is also available. Furthermore we have BLAST4ID tool for ID transfer or annotation. And search as well as download function is accessible. The agriGO can give out rich outputs like graphical result, bar chart result and hierical tree which composing a comprehensive understanding of biological meaning of user's input data.

What is SEA analysis?

SEA analysis means Singular enrichment analysis which is tranditional but widely used. SEA analysis is designed to identify enriched Gene Ontology (GO) terms in a list of microarray probe sets or gene identifiers. Finding enriched GO terms corresponds to finding enriched biological facts, and term enrichment level is judged by comparing query list to a background population from which the query list is derived.

Which statistics method should I choose in SEA tool?

When the input list is compared with the previously computed background, or is a subset of reference list, choose hypergeometric or fisher, for latter only when your query number is quite small. When the input list has few or no intersections with the reference list, the Chi-square tests are more appropriate.

What is PAGE analysis?

PAGE is Parametric Analysis of Gene Set Enrichment [Kim et. 2005 BMC Bioinfomatics]. PAGE method is using Central Limit Theorem in statistics, this method is simple and efficient. Different to SEA, it takes expression level into account, and can deal with a long list of genes/probesets. PAGE use a two-tailed test to count Z score, and the caculation of p-value will be:
if Z score >= 0: p-value is 2 * (1 - x)
if Z score < 0: p-value is 2 * x

What is BLAST4ID?

The BLAST4ID tool is not an analysis tool, but an associated one used mainly for two purposes: 1. Transfer your IDs which are not available to agriGO to available ones, 2. use blast search to annotate your sequences with GO.

Which tool should I choose?

It will depend on what data you have. If you only have a list of identifiers or only interested about them, SEA will be your choice. And if you like take expression data into count and would like compare several dateset then you may try PAGE. The BLAST4ID is only an associated tool, use it if you really need it.

Why graphical/chart image does not display on my PC?

The bar chart result need flash player to browse correctly. And you may need different tool to display different format graphical result, for example: Adobe reader, SVG brower. Contact me if you install related tool but still can not see the results.

How many datatypes are supported by agriGO?

We currently support on 35 species including 280 datatypes. Please check the data statistics page for detail information. We will continue adding more species and datatypes.

How agriGO obtains its data source?

Raw GO annotation data is generated using BLAST, Pfam, InterproScan by agriGO or obtained from B2G-FAR center or from Gene Ontology. Arabidopsis genome data is from TAIR. Rice TIGR genome data is from Rice Genome Annotation Project. Rice KOME data is from KOME database. Rice Gramene data is from Gramene center. Populus genome data is collected from JGI. Soybean and Sorghum genome data is compiled from phytozome. Grape genome data is compiled from Genoscope. Medicago genome data is from Medicago truncatula sequencing resources. Maize genome data is from MaizeSequence.org. Castor bean genome data is from Castor Bean Genome Database. Brachypodium distachyon genome data is from Ensembl. Bovien genome data is from Bovine Genome Database. Silkworm genome data is from SilkDB. M. grisea genome data is from Magnaporthe grisea Database. affymetrixmetrix CSV files and array sequences are from NetAffx.

How often does agriGO update?

Normally we will update our database every 3 months, but if we will update agriGO if some important data source is newly available. Improvement and updating to agriGO tools are irregulated.

Can I check result from old version by new agriGO ?

Sorry, but no. Because we reconstructed the database and redesigned the website organization, analysis result from EasyGO is not supported in agriGO.

How to make agriGO add new customized datatype?

User can contact the agriGO administrator by email (adugduzhou@gmail.com) to discuss more details.